DOCUMENT RESUME

ED 180 980 SP 015 308

AUTHOR Boud, D. J.: Tyree, A. L.
TITLE Self and Peer Assessment in Professional Education: A Preliminary Study in Law. TERC Research and Development Paper No. 55.
INSTITUTION New South Wales Univ., Kensington (Australia). Tertiary Education Research Centre.
PUB DATE Jun 79
NOTE 20p.

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement: *Classroom Participation: Correlation: Discussion (Teaching Technique): *Evaluation Criteria: Foreign Countries: Lawyers: *Legal Education: *Peer Evaluation: Professional Personnel: Research Methodology: Self Concept: *Self Evaluation: Student Attitudes
IDENTIFIERS *Australia (New South Wales)

ABSTRACT

A study of the role of self, peer, and instructor assessment in the self-concept of undergraduate law students focuses specifically on the assessment of students' class participation. The procedure described can be used for the assessment of class participation in any subject. The criteria for assessment generated by students are given, a comparison between self, peer, and teacher marks is made, and the reactions of the students involved in the study are presented. A copy of the instrument used is appended. (LH)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED180980

RESEARCH AND DEVELOPMENT PAPER No. 55

ISSN 0313-5543
ISBN 909528 30 6

SELF AND PEER ASSESSMENT IN PROFESSIONAL EDUCATION: A PRELIMINARY STUDY IN LAW

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. P. Powell
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

D.J. BOUD
Tertiary Education Research Centre
University of New South Wales

A.L. TYREE
Faculty of Law
University of New South Wales

June, 1979

SP 015 308

Tertiary Education Research Centre
University of New South Wales

Self and Peer Assessment in Professional Education:

a preliminary study in law

D.J. BOUD
Tertiary Education Research Centre
University of New South Wales

A.L. TYREE
Faculty of Law
University of New South Wales

Abstract

The importance of assessment by students of themselves (self-assessment) and of their fellow students (peer-assessment) is becoming increasingly recognised especially in professionally oriented courses. This paper explores one way in which such assessments may be conducted within the context of legal education.

The specific focus of the paper is on the assessment of class participation which is a problematic area even in traditional assessment terms. A procedure is described which can be used for the assessment of class participation in any subject. This procedure is applied to a first year undergraduate law class at the University of N.S.W. The criteria for assessment generated by students are described, a comparison between self, peer and teacher marks is made, and the reactions of students are presented.

The results of this experiment are such as to encourage the use of this or similar procedures in other contexts.

Self and Peer Assessment in Professional Education:

a preliminary study in law

I. Introduction

The quality of a lawyer's work is, for the most part, judged by fellow lawyers. This is a position which the legal profession shares with the other professions; it is a position which all of the professions guard jealously, although one which has come increasingly under attack by the wider community in recent years. This "internal" assessment is generally on an informal basis, the criteria and standards used being vague and inexplicit; it is only when something goes terribly wrong that the lawyer is subjected to a formal assessment.

The situation of the professions thus differs markedly from that of many other forms of employment in two ways. First, most assessment is by peers rather than by those in a supervisory capacity. Secondly, the assessment is not usually of such a form that it will tend to improve the day-to-day or long term effectiveness of the lawyer's work; it is of an ill-defined nature which is more related to maintaining the professional image rather than the raising of standards.

This structure of the profession has implications for individual lawyers. The lack of effective external assessment directing day-to-day work means that the ability to define goals and to self-assess progress toward these goals must be better than those of most non-professionals. Lack of this self-assessing ability may eventually produce effects which will meet with adverse assessment of the lawyer by his peers, but such peer-assessment may be too late to have the necessary corrective effect. In the extreme, a lack of ability to self-assess performance will bring into play the formal assessment procedure along with its usually unpleasant consequences.

The ability to self-assess accurately is of further importance to lawyers at the most basic level of practicality: lawyers never have enough time to do all the things they may wish. The ability to self-assess can help a lawyer make optimal use of time available.

If it is accepted that the ability to self-assess is a desirable trait in a lawyer, we must ask what is the responsibility of those in legal education to develop and stimulate that skill in law students. Is it possible to teach self-assessment skills? If it is, should it be a clearly defined aim of our legal education programme? These are broad questions for which we will not attempt definitive answers; but we will argue below that some of the legal skills that we attempt to develop in students have a component of self-assessment and that a clear appreciation of this could help us to sharpen the definition of the objectives of early legal education courses.

This discussion would be entirely superfluous if it were the case that existing law students were perfect self- and peer-assessors. Strangely enough, we know of little previous study of this question, although there has been some activity in the medical sciences (1); the discussion there has progressed to the consideration of self- and peer-assessment procedures as an important adjunct to inadequate policing of the professions (2). In the case of law, a first stage in examining law students' ability to self- and peer-assess is to establish situations in which we can observe students practising these skills. This provided our motivation for the present study.

II. Self assessment skills in today's legal education

The objectives of first year courses in law are usually stated in somewhat general terms. An introductory course will be attempting to develop "lawyer's skills" and to provide some insight into the "common law method". If, as is the current fashion, the introductory course is taught using the case method (or 'Socratic' method), there will probably be an emphasis upon the development of a number of more precisely definable skills.

Emphasis upon the critical reading of case law is an example in point. When first assigned to read a case, the ordinary student reads it through quickly and uncritically; such a student will be surprised (and possibly angered) to find that he or she is unable to successfully answer or explain the questions put by the teacher. When reading the next assignment, he or she will put some of the same sort of questions to him - or her self; as the course progresses, the student should anticipate more and more of the teacher's questions. If this anticipation does occur the student has developed a limited skill of self-assessment in a certain sphere of his or her activity.

Development of self-assessment is further apparent in the teaching of other skills at the beginning stages of a legal education. The teacher emphases the identification of relevant facts and issues, probing and encouraging self-criticism of legal and logical arguments. The student can scarcely be successful at learning these skills without developing an ability to assess his or her own progress within the scope of the skill being learned. It is important to notice that the self-assessment skill does develop.

But it is not only in the area of skills that self-assessment develops in the successful student. We encourage and demand that students begin to identify the policy behind the law. They must think about why the law has developed in the manner in which it has and they must consider the competing interests which must be balanced in the implementation of new laws. Again it will be noticed that the ability to assess his or her own progress is one which develops with practice.

Thus we see that the self-assessment skill is one which, in a limited and inarticulate fashion, we already "teach" in our law schools. It is at least arguable that we should consider the teaching of self-assessment skills as an explicit and well defined objective of our courses: this wider view is certainly sustainable if the role of self-assessment in the professional lives of lawyers is acknowledged.

III. Current Ability at Self- and Peer-Assessment: A Case Study

Given the importance of self- and peer-assessment for lawyers, it is somehwat surprising to find an almost total absence of research on the question of law students' ability at assessment. The only example that could be found was a study in the U.S.A. on the impact of self-grading on performance and evaluation in a constitutional law course in a political science department (3).

We decided to investigate students' ability to make assessments of themselves and their peers in a one-session (15 week) first year class entitled "The Legal System" taught by the second author at the University of New South Wales in 1978. The class size was 28 and was taught in the 'Socratic' fashion. The course was structured around the development of consumer protection through the common law process of implying terms into express contracts of sale of goods or provision of services. Aims of the course included the development of analytical skills and case law technique.

Admission to the Law course at New South Wales is restricted; students are admitted on the basis of their marks in a State-wide examination (the Higher School Certificate). There are, in addition, a few places allotted to mature-age students and a special programme for the admission of Aboriginal students. One of each of these special categories participated in this experiment.

The experiment was limited to the assessment of class participation. This is an important part of the assessment process in law at New South Wales and has been the subject of two previous studies (4). Class participation accounted for a large part of the overall mark in the Legal System course, being 20% of the final mark for the course.

As a means of examining the students' behaviour in self- and peer-assessment, class participation has other appealing characteristics. Although it is a relatively small part of the overall contribution of a student to the course, it is a very public part. There was no need for the students to read the work of others or, indeed, to prepare in any special manner for the exercise.

Two aspects of students' ability to make assessments were considered:

(a) establishing appropriate criteria for assessments; and,

(b) rating by students of themselves and their peers in terms of these criteria.

The setting-up of criteria for making judgements was just as important as making the judgements themselves. Students would have to think carefully about the aims of the course and the role of class participation in the course before determining the criteria they would use.

The procedure used was that a series of instructions was provided for each student. These gave some guidelines for constructing assessment criteria, and focussed the attention of students on the need to establish priorities and the need to have clearly written criteria. Once the students had constructed the criteria they would use they made an assessment of both their own performance in class participation and that of the other students in the class.

The choice of a suitable scale to use for rating was the subject of some debate between the authors. The simplest method was to use a ten-point scale with students giving a mark out of ten on each criterion. This was familiar and easy to use but it suffered from two main drawbacks. First, a mark of, say, five has certain connotations independent of the criteria being used; that is, it is regarded as a bare pass and the student receiving it may be thought to have bare-pass qualities. Secondly, there is the possibility of the entire class using a very small band of marks. This would lead to a lack of discrimination between students, causing difficulty in interpreting the results of the exercise, or possibly to the inflation of grades, as has been reported elsewhere (5).

An alternative to the ten-point scale was to use what is called a self-normalising scale. Using such a scale marks are allocated so that the full range is used and that equal numbers of students appear above and below the average for the class. This scale has two advantages. First, no grade inflation can occur because no matter which students are involved the average for the class is defined in advance. Secondly, it is more valid to compare the marks of different students as there will be less variation in the method of marking. The main disadvantage is the strangeness of the method. Most students and teachers know, or think they know, how to use a ten-point scale even though they might not use it in an identical fashion, but when the method is substantially varied they need special instructions.

As part of the experiment it was decided to use both types of scale and to compare the effects. Once students had established the criteria they were to use they were asked to mark themselves and their peers using Method A: the scale 1,2,3,4,5,6,7,8,9 with 5 as the pass mark; and Method B: the scale -2,-1, 0,+1,+2 with the rule that there should be equal numbers of students above and below the mean of 0. Method B thus approximates a normalised scale.

A shortened version of the guidelines provided for the students is given in the Appendix.

The sequence was as follows:

1. Each student listed all the criteria which he or she thought were important in assessing class participation.
2. They clarified these and elaborated on them as necessary.
3. They selected the three criteria which they considered to be the most important.
4. They rated themselves and other members of the class using each one of these three criteria in turn on both Method A and Method B.

At the end of the exercise, which took approximately one hour of class time to complete, the following marks were available for each student on each criterion using both methods:

Self-mark (mark given by himself or herself)

Peer-mark (Average of marks, given by the other students)

Teacher-mark (mark given by the teacher)

Results

Twenty-five students were present in the class when the experiment was conducted. Although all students had been told about it in advance they did not know at what time it would take place. Ratings were obtained from all twenty-five students for Method A and from twenty-three for Method B. Two students had completely misunderstood the instructions for Method B and their results could not be used, and eleven others had not adhered to the rule of equal numbers above and below the mean. Some of these students tried to justify their non-observance of this rule, one commenting, "everyone should consider themselves 'above average'"!

Criteria

The students produced a total list of 148 criteria, an average of 5.9 per student with a range from 5 to 8. These criteria were subsequently classified under the following headings:

1. Characteristics of an individual's contribution

   Cognitive: logic, objectivity, knowledge, creativity.

   Expressive: clarity, fluency, conciseness.

   Affective: enthusiasm, interest.

   Contribution to the process of learning: recognition of the responses of others, constructive criticism, contribution to group climate, relevance.

2. Inferred preparation: Amount, consistency, timeliness.
3. Attendance.

This part of the exercise was received very favourably by the students. One student commented:

> This is an excellent exercise, for it makes one aware of what a good student should do.

Another wrote:

> These steps allowed those of us involved to suggest whichever factors we wished and express them in whatever way seemed to us the most appropriate. This lack of rigidity is excellent and may allow us to present to staff factors for judging performance which previously they had not considered.

The latter comment was quite prescient. The factors which were outlined by the students were all relevant in the view of the teacher and provided a considerably broader list than the one that he initially generated.

The main dissatisfaction with this aspect of the exercise was that having specified a series of important criteria they only had an opportunity for three of these to be used for assessment. This view was expected, and was considered by the authors to be an inevitable consequence of the need to limit the exercise to a manageable scale.

A significant criticism was voiced by another student:

> I think the factors would be better arrived at by class discussion because one person is liable to miss out on an important one, or word it poorly. Once decided on, they should all be used.

If a procedure similar to the one we have used under experimental conditions were used for student assessment then considerations of parity and fairness would provide strong arguments for a common set of criteria. However, there is a limitation in this for, no matter how well worded a given criterion might be, it would always be subject to multiple interpretations, and thus an apparently common scale would, in practice, not be assessing the common factors for which it was designed.

Ratings

For convenience in presenting the results, we show only the mark obtained by averaging over the marks on the three criteria. Scatter plots for each of the comparisons may be found in Appendix II.

Comparisons between the marks given to each student by themselves, their peers, and their teacher are shown in Table 1. Product-moment correlation coefficients were calculated for each comparison for both methods. The reader should consult Appendix III for comments concerning statistical significance.

The main features of these results are as follows:

1. Students tend to rate themselves more favourably than they are rated by their peers.

2. Students rate themselves less favourably than they are rated by the teacher using Method A, but this is completely reversed when Method B is used.

3. Peers tend to be harsher in their ratings than the teacher when Method A is used, but using Method B this is reversed and the peers are somewhat more lenient than the teacher.

4. In general on both methods there is a very high level of agreement between the marks given by peers and those given by the teacher.

Table 1

Comparison of self- and teacher- ratings

	Method A	Method B
Number of students with marks:		
self- higher than teacher-	9	17
self- lower than teacher-	16	6
	25	23

r_A = 0.46* r_B = 0.55**

Comparison of peer- and teacher- ratings

	Method A	Method B
Number of students with marks:		
peer- higher than teacher-	4	16
peer- lower than teacher-	21	9
	25	25

r_A = 0.75*** r_B = 0.83***

Comparison of self- and peer- ratings

	Method A	Method B
Number of students with marks:		
self- higher than peer-	17	16
self- lower than peer-	8	7
	25	23

r_A = 0.51** r_B = 0.60**

Table 1. Comparisons of student marks: The stars indicate the statistical significance of each correlation. (* = significant at .05 level, ** = significant at .01 level, *** = significant at .001 level).

One of the difficulties of comparing sets of marks is that different people tend to use different parts of the same scale. It is quite common for different tutors to give the same piece of work widely differing assessments. A statistical device can be used to modify sets of marks so that different markers appear to produce the same mean mark and the same spread of marks. This then enables direct comparisons to be made between the marks received by a particular student from different markers. The procedure is known as producing a normalised score.

Now if the students had correctly interpreted our instructions for Method B the calculation of the normalised B scores should leave their results effectively unchanged, for what the normalising procedure does after the marks have been given is essentially the same as what the 'self-normalising' Method B aimed to do.

Table 2 presents the same comparisons as Table 1, but in terms of the normalised scores. The correlations are unchanged as this procedure has no effect on them. The fact that the results of normalising the B score does show a considerable shift is evidence of the degree to which students failed to follow our instructions.

Table 2

Comparison of self- and teacher- ratings	Method A	Method B
Number of students with marks:		
self- higher than teacher	10	9
self- lower than teacher	14	14
	25	23

Comparison of peer- and teacher- ratings	Method B	Method B
Number of students with marks:		
peer- higher than teacher-	15	11
peer- lower than teacher-	10	14
	25	25

Comparison of self- and peer- ratings	Method A	Method B
Number of students with marks:		
self- higher than peer-	14	14
self- lower than peer-	11	9
	25	23

Table 2. Comparisons of student marks using normalised scores.

When the normalised results are considered the findings become clearer:

1. Students tend to rate themselves more favourably than they are rated by their peers.

2. Students tend to rate themselves less favourably than they are rated by their teacher.

Our other finding remains unchanged:

3. There is a very high level of agreement (i.e., correlation) between the marks given by peers and those given by the teacher.

The differences between the distributions of the results using the two methods is an important finding, and, if it is duplicated in further studies, could lead to a reassessment of the previous work that has been done on self- and peer-assessment. If it does hold elsewhere it indicates that the choice of a scale is an extremely important issue when comparisons are to be made between the assessments generated by different groups. If the finding that students rate themselves more or less harshly than the teacher can so easily be reversed then the acceptability of self-assessment could be simply influenced by the choice of an appropriate scale.

Student preferences for Method A or Method B were very polarised. They seemed to appreciate the issues involved in using the different scale, but plumped for one or the other. Nine students gave a clear preference for Method A and six for Method B, but some of these appeared to be voting against the alternative. For example, one student wrote of Method B:

> Horrible! It is quite unfair to give someone -2 simply because someone else has scored +2. The mathematical logic of this section is not appropriate when applying to classes.

This student is making the point that there are more issues to be considered when making a choice of a rating scale than simply the logical ones; students disliked being forced to give negative scores. Assessment is an emotive subject and procedures need not only to be scrupulously fair, but they must also be felt by students to be fair. Marks have, unfortunately, an association with status and nobody seems to want changes in status for purely logical reasons.

IV Conclusions

The most gratifying results of the experiment were the positive responses from the students. They felt that they had participated in a worthwhile learning experience; our reading of the criteria which they constructed for the purpose of assessment leads us to the same belief. If our opening comments concerning the role of self- assessment in the profession are accepted, there is a good case for extending this learning experience throughout the law curriculum.

We have already discussed the numerical results of the experiment. We note that the correlation between each pair of assessors is good, but that the students seem to have used a different part of the "standard" scale than did the teacher. It was in anticipation of this problem that we introduced the "self-normalising" scale. However, a significant proportion of the students found this scale to be

objectionable. Using normalised scores from the "standard" scale for the purposes of comparison would seem at least as satisfactory as attempting to construct a more sophisticated method of marking.

As always, a study of this sort raises more questions for consideration. The results of the students' participation were not a part of the formal assessment of the course, a fact which was known to the students at the time of the exercise. Would it make any difference to the comparative results? Furthermore, it might be hoped that more advanced students would have gained greater facility in the self-assessment process, even if by "greater facility" we mean only that the students have a more accurate assessment of the teacher's expectations.

We believe that the questions raised in the preceding paragraph are not independent, but may be influenced by an underlying factor, We conjecture that in a climate of greater competitive tension the agreement between the self/peer/teacher assessment could be expected to deteriorate as self-preservation instincts come to the foreground as conscious or unconscious motives in the assessment procedure. If this conjecture is correct, then including the experiment as part of the formal assessment procedure would decrease the observed agreement.

But a decrease in the observed agreement should not, in itself, be grounds for rejecting the use of self and peer assessment as a part of the formal assessment process. If the educational benefits of such an exercise are great enough, we should be prepared to modify the assessment procedure in such a way that it is compatible with our educational goals.

REFERENCES

1. Kennell J.H. et al (1973) Self evaluation by first year medical students in a clinical science programme British Journal of Medical Education 7, 4 pp 230-238; Morton J.B. and Macbeth W.A.A.G (1977) Correlations between staff, peer and self assessments of fourth year students in surgery Medical Education 11, 3 pp 167-170.

2. Sanazaro P.J. et al (1974) Medical Audit British Medical Journal 1, 271

3. Stover R.V. (1976) The impact of self-grading on performance and evaluation in a constitutional law course Teaching Political Science, 3, 3 pp 303-310.

4. Armstrong M. (1978) Assessing Students' Participation in Class Discussion Assessment in Higher Education 3, 3 pp 186-202; Hayes S.L. and Hayes R.A. (1973) Towards objective Assessment of Class Participation Journal of the Society of Public Teachers of Law, 12 pp 323-332.

5. Huntley J.F. (1976) Academic Evaluation and Grading; an analysis and some proposals Harvard Educational Review 46, 4 pp 612-631.

6. Freund J.E. Mathematical Statistics, 1962, Englewood Cliffs.

Appendix I: Guidelines for Students

Developing Awareness of Learning

One purpose of an assessment procedure is to help you monitor your progress in the subject. The current assessment schemes depend upon criteria which are defined by someone other than yourself, usually the teacher. We feel that it would be very advantageous for you to be able to devise your own assessment programme; if this is done habitually in your professional life, you will find that your development will be much more rapid as well as being directed toward goals that you believe to be important. Our goal is to provide you with tools to answer the question "how am I doing?".

This exercise is a very modest step toward the above goal. It is intended only to apply to the question of contributions made in class; other important areas of assessment are outside the scope of this exercise because of time limitations, but we hope that you will attempt to make use of the methods of this exercise in other areas of your work.

The task is in two parts:

i. finding and describing the criteria that you would use to assess contributions to the class;

ii. applying these criteria to assess your own contributions and those of your colleagues.

We hope that there will be two worthwhile results from this exercise. First, that you will increase your awareness of criteria for good work in this subject and in every task which you undertake. Second, that we may find a method of using these forms of student-generated-assessment in a formal assessment procedure. The results of this exercise will not from part of your assessment for this course.

Work through each step of the following sequence in order, writing in the spaces provided where that is appropriate. Do not miss out any steps. If you wish to comment please do so on the sheet provided on the last page or write your comments clearly in the margin.

1. What factors would you take into account when judging the worth of other students' contribution to the course? Spend about 5 minutes thinking about this before writing anything down. Include everything that you can think of (at least five factors). For example, "knowledge of set readings". "clarity of exposition". etc.

2. What factors would you take into account when judging the worth of your own contributions to the course? Are there any additional factors to those listed above? If there are please list these. Also, if you want to reword any of those above to make them relevant here, please do so.

3. Look carefully at the two lists above and go through each item deciding whether or not it is possible for you to observe the extent to which students can notice them in each other. For example, you may have listed "quality of written work" in list 1. but other students may never get a chance to see this, so you would delete that from your list. Put a line through each deleted item in such a way that you do not obscure what is underneath.

4. Consider the items that are left. Are they worded in such a way that they are clearly understood by other people? Sometimes what is clear to you is ambiguous to others. Do you need to explain any of your items to make them clear? If you would, or if you can state any of them more clearly, do so now. Rewrite your clarified list below. Taking the example above, you might add "knowledge of set readings exhibited in class discussions".

5. Now decide what priorities you would give to your items. Which are the most important, which are the least important? When you are making judgements in practice you will need to limit yourself to simplify your task. Number them in order of priority: 1,2,3, ...

6. Select the three at the top of the list. Write them out again in the space provided on the next sheet. For the purpose of this exercise we shall limit ourselves to the top three.

7. The next stage is to rate yourself and others according to the extent to which you have contributed to the course using the factors you have just identified. There are at least two ways to do this. The first is easy and corresponds to the normal marking schemes you will have met at school and university. The second is more complicated but it does have some technical advantages which will be described to you later if you wish (it would take up too much space here). We would like you to do both so that you (and we) get a chance to compare them.

(The remainder of the guidelines included details of the two rating methods and blank mark sheets with students' names.)

Appendix II : Comparisons of self-, peer-, and teacher marks

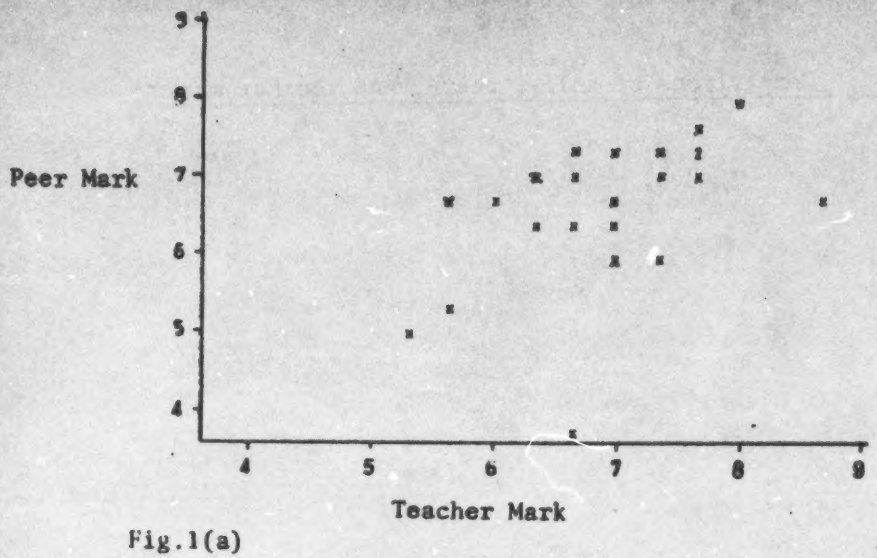

Fig.1(a)

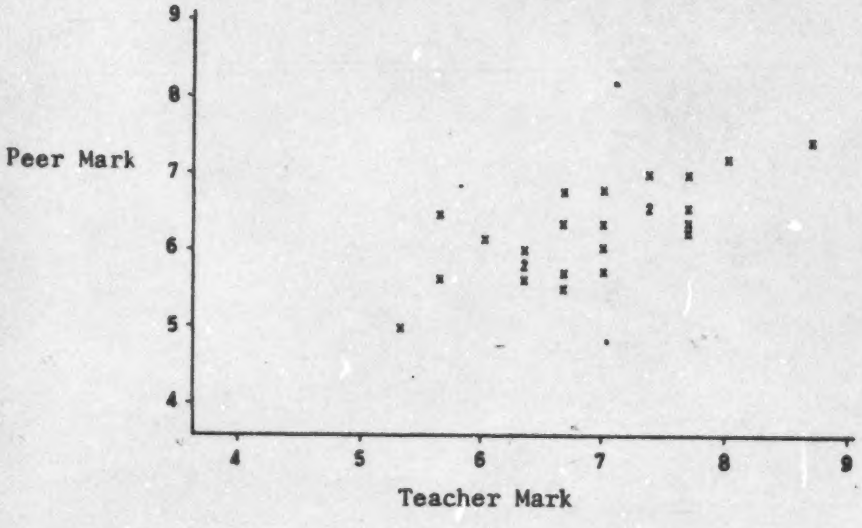

Fig 1(b)

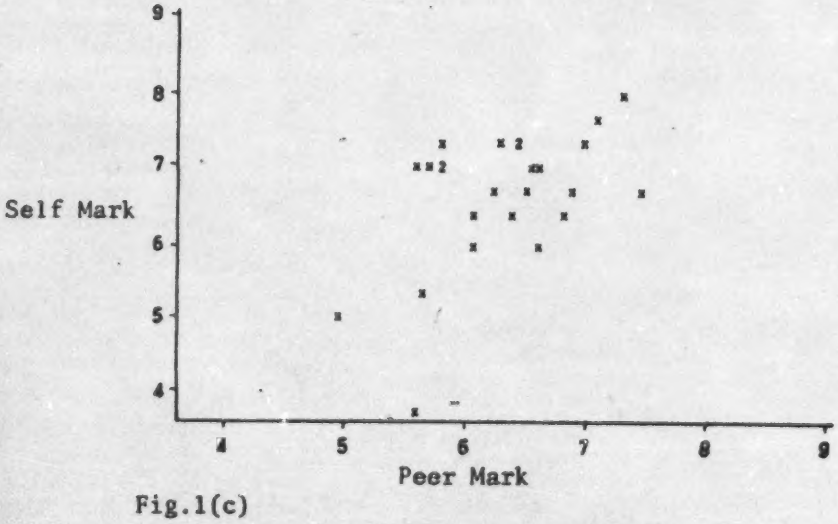

Fig.1(c)

mparison of Marks using Method A

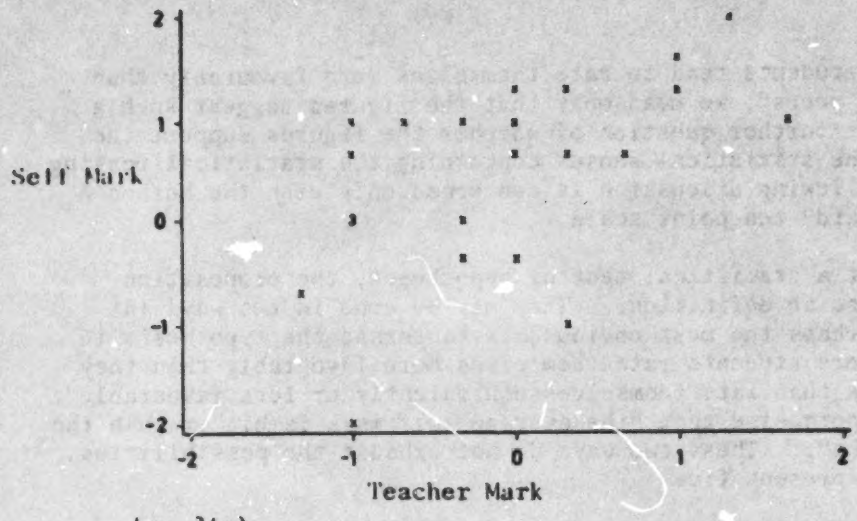

Fig.2(a)

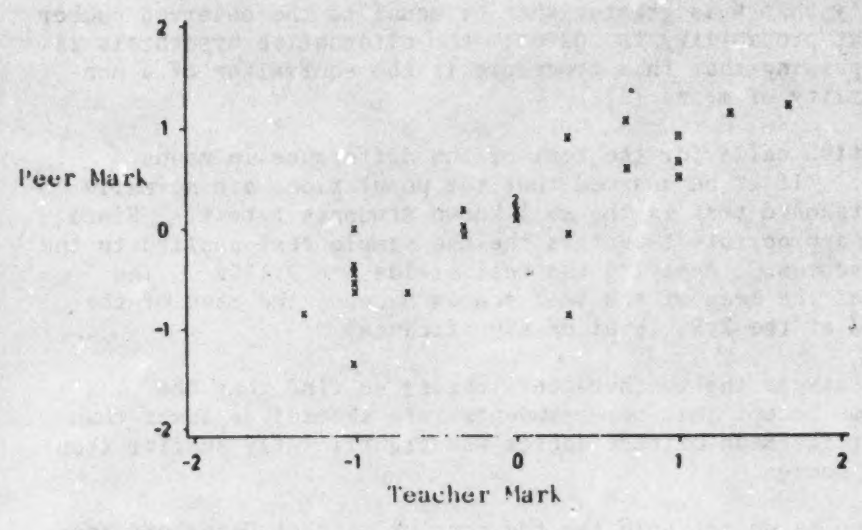

Fig.2 (b)

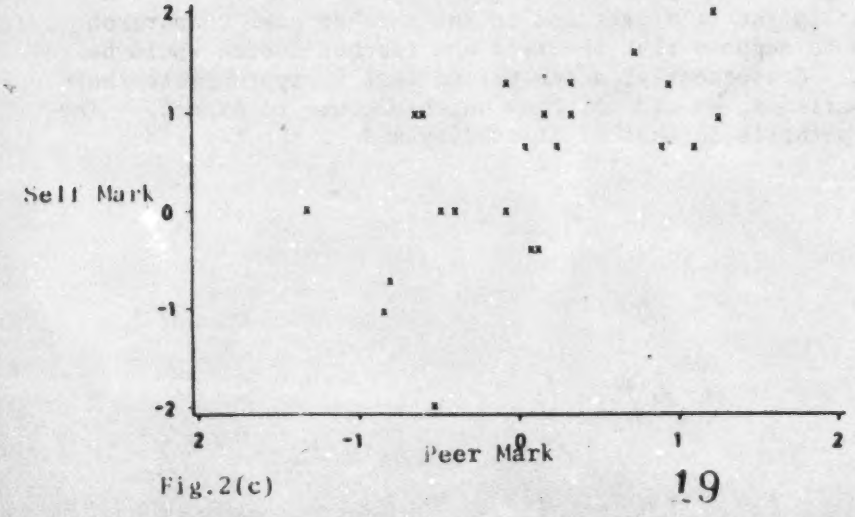

Fig.2(c)

Comparison of Marks using Method B.

Appendix III : Statistical Significance

When we say that "students tend to rate themselves more favourably than they are rated by their peers", we mean only that the figures suggest such a tendency. There is the further question of whether the figures support the proposition in any of the statistical senses concerning the statistical testing of hypotheses. The following discussion is concerned only with the Method A scores, i.e. the "standard" ten point scale.

In order to conduct a statistical test of hypotheses, the proposition must be given a more precise definition. This may be done in two ways (at least). The first, perhaps the most obvious, is to phrase the hypothesis in a form equivalent to "more students rate themselves more favourably than they are rated by their peers than rate themselves equivalently or less favourably". The second way is to hypothesise that "the average self mark is higher than the average of the peer marks". These two ways do not exhaust the possibilities, but will suffice at the present time.

Taking the first alternative, we note that 17 of the students rated themselves more highly than they were rated by their peers. If we assume a Bernoulli Model with P not exceeding 1/2, then the number of successes K is Binomially distributed. To test $P = 1/2$ against the alternative $P > 1/2$, is to calculate the probability that K is greater than or equal to the observed number of "successes" 17. That probability is .02164; the alternative hypothesis is accepted. We note in passing that this procedure is the equivalent of a non-parametric test for equality of means (6).

The second alternative calls for the test of the difference in means between two populations. If it be assumed that the populations are normally distributed, then the standard test is the well-known Students t-test. Since the data is paired, the appropriate t-test is the one sample test applied to the differences of the two scores. Applying the test yields $t = 2.1158$. The research hypothesis, that the mean of the self scores exceeds the mean of the peer scores, is accepted at the 2.5% level of significance.

Applying similar tests to the teacher-peer scores, we find that the situation is similar; we accept that more students rate themselves lower than did the teacher and that the mean of peer scores was significantly smaller than the mean of the teacher scores.

The statistical results do not hold for the conclusion that "students tend to rate themselves less favourably than they are rated by the teacher". Thus this conclusion must be regarded with considerably more caution than our other results.

We further note that in the self/peer and in the teacher/peer comparison that we had prior reason to suppose that the self and teacher scores would be higher than peer scores. Consequently, a one-tailed test is appropriate there. In making the other comparisons, we did not know which outcome to expect. The appropriate alternate hypothesis is that of inequality and so the two sided critical region was used.